

Requested Patent: JP4330563

STL 9-2000-0077-US4

Title: CONNECTION VALUE STRUCTURE ANALYSIS SYSTEM

Abstracted Patent: JP4330563

Publication Date: 1992-11-18

Inventor(s): KOMATSU JUNKO

Applicant(s): RICOH CO LTD

Application Number: JP19910130627 19910502

Priority Number(s):

IPC Classification: G06F15/38 ; G10L3/00

Equivalents:

ABSTRACT:

PURPOSE: To easily generate a practical connection value dictionary and semantic classification compatible with this dictionary with a high precision.

CONSTITUTION: A connection value pattern generating part 2 collects modification examples of each declinable word from modification examples stored in a modification example data base 1 to generate a connection value pattern, and the connection value pattern generated in this stage is verified in consideration of compatibility with the whole of modification examples to obtain the final connection value pattern. A semantic classification generating part 3 regards words (nouns) belonging to the same connection value slot in the connection value pattern, which is obtained by the connection value pattern generating part 2, as words belonging to a resembling word group to generate the semantic classification and resembling word groups are mutually collected in this stage to obtain the final semantic classification.

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平4-330563

(43) 公開日 平成4年(1992)11月18日

(51) Int.Cl. ⁵	識別記号	片内整理番号	F I	技術表示箇所
G 0 6 F 15/38	J	9194-5L		
G 1 0 L 3/00	H	8946-5H		

審査請求 未請求 請求項の数3(全4頁)

(21) 出願番号 特願平3-130627

(22) 出願日 平成3年(1991)5月2日

(71) 出願人 000006747

株式会社リコー

東京都大田区中馬込1丁目3番6号

(72) 発明者 小松 順子

東京都大田区中馬込1丁目3番6号 株式会社リコー内

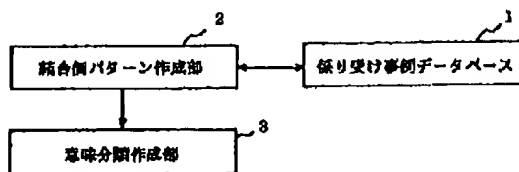
(74) 代理人 弁理士 植本 雅治

(54) 【発明の名称】 結合価構造解析方式

(57) 【要約】

【目的】 実用性のある結合価辞書およびそれと整合性のとれた意味分類を容易にかつ精度良く作成することが可能である。

【構成】 結合価パターン作成部2は、係り受け事例データベース1に記憶されている係り受け事例から各用言毎に類似の係り受け事例をまとめて結合価パターンを作成し、さらにこの段階で作成された結合価パターンを係り受け事例全体との整合性を考慮して検証して最終的な結合価パターンを得る。また、意味分類作成部3は、結合価パターン作成部2により得られた結合価パターンにおいて、同一の結合価スロットに属する単語(名詞)を類似単語グループのものとみなして意味分類を作成し、さらにこの段階で類似単語グループ同士をまとめて最終的な意味分類を得る。



(2)

特開平4-330563

1

【特許請求の範囲】

【請求項1】 係り受け事例を記憶する係り受け事例記憶手段と、係り受け事例記憶手段に記憶されている係り受け事例をまとめて結合価パターンを作成する結合価パターン作成手段と、結合価パターン作成手段により得られた結果を基に名詞の意味分類を作成する意味分類作成手段とを備えていることを特徴とする結合価構造解析方式。

【請求項2】 前記結合価パターン作成手段は、係り受け事例記憶手段に記憶されている係り受け事例から各用言毎に類似の係り受け事例をまとめて結合価パターンを作成し、この段階で得られた結合価パターンを係り受け事例全体との整合性を考慮して検証し、最終的な結合価パターンを作成するようになっていないことを特徴とする請求項1記載の結合価構造解析方式。

【請求項3】 前記意味分類作成手段は、結合価パターン作成手段によって作成された結合価パターンから得られる類似単語グループをさらにまとめて意味分類を作成するようになっていないことを特徴とする請求項1記載の結合価構造解析方式。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、日本語ワードプロセッサにおける仮名漢字変換やテキスト音声合成システムの言語解析等の種々の自然言語解析に利用され、動詞などの用言のもつ結合価パターン構造をベースとする結合価構造解析方式に関する。

【0002】

【従来の技術】 一般に、種々の自然言語解析においては、解析結果が一意に定まらない場合が多い。例えば、日本語ワードプロセッサにおける仮名漢字変換では、同音語（読みが同じで表記が異なる単語）があるために、また、テキスト音声合成システムの言語解析では、同形語（表記が同じで読みが異なる単語）があるために、その解析結果が一定に定まらないことがある。

【0003】 このような曖昧性を解消するために、従来、例えば1987年7月24日発行の文献「自然言語処理 62-6 第39頁乃至第44頁」に開示されているように、動詞などの用言とその支配要素である格要素との結び付きを記述した結合価パターンと呼ぶ構造パターンをベースとして、解析を行なう結合価構造解析方式が知られている。

【0004】 この種の結合価構造解析方式においては、それぞれの用言のもつ結合価構造を適当な枠組みに形式化し、結合価パターンとして結合価辞書に記述し、また、格助詞がとりうる名詞に対し意味分類を作成するようになっている。

【0005】

【発明が解決しようとする課題】 ところで、上述したような従来の結合価構造解析方式において、結合価辞

2

書、すなわち結合価パターンの作成および意味分類の作成は、通常、人手を介して行なわれている。しかしながら、自然言語解析等では実際に利用する単語数が多く、また同じ単語でも文脈や使用分野等によって様々な意味をもつため、実用性のある結合価辞書およびそれと整合性のとれた意味分類を人手によって作成するのは非常に困難であり、従って、従来では、実用性のある結合価辞書およびそれと整合性のとれた意味分類を容易にかつ精度良く作成することができないという欠点があった。

【0006】 本発明は、実用性のある結合価辞書およびそれと整合性のとれた意味分類を容易にかつ精度良く作成することが可能な結合価構造解析方式を提供することを目的としている。

【0007】

【課題を解決するための手段】 上記目的を達成するために本発明は、係り受け事例を記憶する係り受け事例記憶手段と、係り受け事例記憶手段に記憶されている係り受け事例をまとめて結合価パターンを作成する結合価パターン作成手段と、結合価パターン作成手段により得られた結果を基に名詞の意味分類を作成する意味分類作成手段とを備えていることを特徴としている。

【0008】 また、結合価パターン作成手段は、係り受け事例記憶手段に記憶されている係り受け事例から各用言毎に類似の係り受け事例をまとめて結合価パターンを作成し、この段階で得られた結合価パターンを係り受け事例全体との整合性を考慮して検証し、最終的な結合価パターンを作成するようになっていないことを特徴としている。

【0009】 また、意味分類作成手段は、結合価パターン作成手段で作成された結合価パターンから得られる類似単語グループをさらにまとめて意味分類を作成するようになっていないことを特徴としている。

【0010】

【作用】 本発明では、係り受け事例記憶手段に記憶されている係り受け事例をまとめて結合価パターン、すなわち結合価辞書を作成し、作成された結合価パターンを基に名詞の意味分類を作成する。

【0011】 結合価パターンを作成する際には、係り受け事例記憶手段に記憶されている係り受け事例から各用言毎に類似の係り受け事例をまとめて結合価パターンを作成し、この段階で得られた結合価パターンを係り受け事例全体との整合性を考慮して検証し、最終的な結合価パターンを作成するようになっている。

【0012】 また、意味分類を作成する際には、結合価パターンから得られる類似単語グループをさらにまとめて意味分類を作成するようになっている。

【0013】

【実施例】 以下、本発明の一実施例を図面に基づいて説明する。図1は本発明の一実施例のブロック図であり、本実施例では、係り受け事例を記憶する係り受け事例デ

(3)

特開平4-330563

3

データベース1と、係り受け事例から結合価パターンを作成する結合価パターン作成部2と、結合価パターン作成部2により得られた結果を基に、名詞の意味分類を作成する意味分類作成部3とが設けられている。

【0014】上記結合価パターン作成部2は、基本的には係り受け事例データベース1に記憶されている係り受け事例から各用語毎に類似の係り受け事例をまとめて結合価パターンを作成するが、さらにこの段階で作成された結合価パターンを係り受け事例全体との整合性を考慮して検証し、最終的な結合価パターンを得るようになっている。

【0015】また、意味分類作成部3は、結合価パターン作成部2により得られた結合価パターンにおいて、基本的には、同一の結合価スロットに属する単語（名詞）を類似単語グループのものとみなして意味分類を作成するが、さらにこの段階で類似単語グループ同士をさらにまとめて最終的な意味分類を得るようになっている。

【0016】次にこのような構成における結合価構造解析処理について説明する。

【0017】係り受け事例データベース1に、いま例えば図2に示すような形で係り受け事例（A）乃至（F）が記憶されているとする。結合価パターン作成部2では、各用語毎に類似の係り受け事例をまとめる。例えば、用語を共通にする係り受け事例間の類似度Dを、次式のように定義する。

【0018】

【数1】

$$D(a, b) = \frac{\sum_{i=1}^n w_i \cdot f_i(a, b)}{\max(D(a, a), D(b, b))}$$

【0019】このように定義した後、 $D > TH$ の関係を満たすデータを集めて、1つの結合価パターンとして作成する。なお、数1において、 w_i は格助詞1に対する重みであり、 $f_i(a, b)$ は格助詞1がとる名詞の類似度であり、また、上記THは所定の閾値である。名詞の類似度 $f_i(a, b)$ は、例えば格助詞1がとる名詞が同一であるときには“1”の値に、格助詞1がとる名詞が類似しているときには t_1 の値に、 a, b 共に格助詞1が付くときには t_2 の値に、上記以外の場合は“0”の値に設定される。ここで、 t_1, t_2 は、次式の関係にあるものとする。

【0020】

【数2】

$$0 < t_2 < t_1 < 0$$

【0021】係り受けデータベース1に図2に示すような係り受け事例（A）乃至（F）が記憶されているときには、結合価パターン作成部2は、これらの係り受け事例（A）乃至（F）を上記数1を用いてまとめ上げ、図3（a）、（b）、（c）に示すような3つの結合価パ

4

ターンとして作成する。

【0022】通常、この段階において、同一の結合価スロットにくる名詞にはかなりの類似性がみられ、実際、図3（a）、（b）の結合価パターンにおいて、同一の結合価スロットにくる名詞、例えば「信用」と「信頼」とは、類似の意味を持つ。しかしながら、図3（c）の結合価パターンでは、動詞「かける」に複数の意味があるので、同一の結合価スロットにくる名詞、「水」と「時間」とは、類似の意味をもたず、これらは、本来、同一の結合価スロットに属させるべきではない。すなわち、同一の結合価スロットに属させると、「水」と「時間」を同じ意味にしてしまうような誤りが生じる恐れがある。

【0023】そこで、結合価パターン作成部2は、図3（a）、（b）、（c）のように結合価パターンを作成したときに、この段階で、さらに、作成された結合価パターンを係り受け事例全体との整合性を考慮して検証する。例えば、作成された結合価パターンにおいて、類似の意味をもつと仮定された同一の結合価スロットにくる名詞が本当に類似の意味をもつものであるかを検証する。この検証に際し、類似の意味をもつ名詞は、同じ使われ方をされる可能性が高いという仮定に基づいて、結合価スロットSに属する名詞Naが結合価スロットSに属する他の名詞と類似しているもつもらしさCを次式のように定義する。

【0024】

【数3】

$$C = \frac{\text{和}(N(Na, Ni))}{N(Na)}$$

【0025】ここで、 Ni は結合価スロットSに属するNa以外の名詞であり、和 $(N(Na, Ni))$ はNaと Ni が共に属する結合価スロットの種類であり、また、 $N(Na)$ は係り受け事例中のNaの出現頻度である。

【0026】結合価パターン作成部2は、数3に基づき検証を行ない、Cが極端に小さい場合には、Naが結合価スロットに属しているのは誤りである可能性が高いと判断し、naを結合価スロットSから削除する。このようにして、誤りのない正しい結合価パターンの記述を結合価辞書として最終的に得ることができる。

【0027】次いで、意味分類作成部3では、結合価パターン作成部2によって作成された結合価パターンにおいて、先づ、同一の結合価スロットに属する単語を類似単語グループのものとみなし、これに同一の意味分類コードを付加する。例えば、結合価パターン作成部2による処理の結果、最終的に、図4（a）、（b）に示す結合価パターンだけが残ったとすると、意味分類作成部3は先づ、「信用」と「信頼」とにコード“1”を付加し、「首相」と「大統領」とにコード“2”を付加し、「アメリカ」と「ソ連」とにコード“3”を付加する。

(4)

特開平4-330563

5

このような意味分類コードはあくまでも形式的なものであり、意味分類コードが異なっているこれらの意味分類を構成する名詞が互いに類似しているものもある。従って、意味分類作成部3は、上記のような意味分類を行なった後、さらに類似単語グループ同士を例えばクラスタリング法などを用いてまとめ上げ、意味分類を作成する。すなわち、異なる意味分類コードが付加されたもの同士であっても、意味分類を構成する名詞が互いに類似している場合には、これらを1つにまとめ上げて最終的な意味分類を作成する。なお、単に意味分類をまとめるだけでなく、それらの階層関係を保持しておけば、結合価辞書を利用する際に役立てることができる。

【0028】このように本実施例では、係り受け事例データベース1に記憶されている係り受け事例から結合価パターンが自動的に作成され、また、作成された結合価パターンを基に名詞の意味分類が自動的に作成されるので、人手を介さずに結合価辞書およびそれと整合性のとれた意味分類を容易に作成することが可能となる。

【0029】この際に、係り受け事例データベース1をユーザの利用分野に適応した事例によって構成すれば、これに基づきユーザの利用分野に適応した精度の高い実用的な結合価辞書および意味分類を作成することができる。

【0030】

【発明の効果】以上に説明したように本発明によれば、係り受け事例記憶手段に記憶されている係り受け事例をまとめ上げて結合価パターン、すなわち結合価辞書を作成し、作成された結合価パターンを基に名詞の意味分類

6

を作成するようになっているので、実用性のある結合価辞書（結合価パターン）およびそれと整合性のとれた意味分類を人手によらずに容易にかつ正確に作成することができる。

【0031】結合価パターンを作成する際には、係り受け事例記憶手段に記憶されている係り受け事例から各言毎に類似の係り受け事例をまとめ上げて結合価パターンを作成し、この段階で得られた結合価パターンを係り受け事例全体との整合性を考慮して検証し、最終的な結合価パターンを作成するようになっているので、実用性のある結合価辞書を精度良く作成することができる。

【0032】また、意味分類を作成する際には、結合価パターンから得られる類似単語グループをさらにまとめ上げて意味分類を作成するようになっているので、より正確にかつコンパクトに分類された意味分類を作成することができる。

【図面の簡単な説明】

【図1】本発明の一実施例のブロック図である。

【図2】係り受け事例データベースに記憶されている係り受け事例の一例を示す図である。

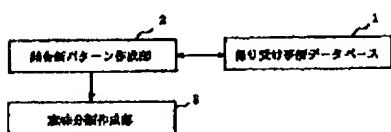
【図3】(a)、(b)、(c)は結合価パターンの一例を示す図である。

【図4】(a)、(b)は最終的に得られた結合価パターンの一例を示す図である。

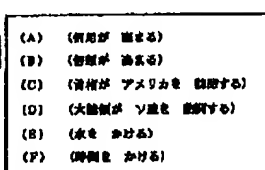
【符号の説明】

- 1 係り受け事例データベース
- 2 結合価パターン作成部
- 3 意味分類作成部

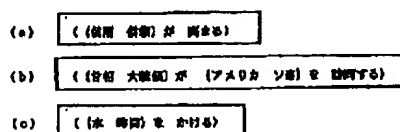
【図1】



【図2】



【図3】



【図4】

